

Part – 1

Objective: Use Spark Map Reduce functions in order to compare friends lists and create a new list of user pairs and their mutual friends

Data: The data consists of approximately **88,000 friend pairs**. Each line is a ‘user-friend’ and a single ‘friend-from-friend-list’, each represented by an id number. So, the data consists of many rows similar to ‘0 1’ or ‘176 1130’ with the first number being the ‘user’ and the second number being one of their friends.

However, I was unable to get this data format to work and had to create my own.

```
0 1 2 4 5 6 7
1 0 2 3 5 9
2 1 0 6 7 8 9
3 1 4 5 7 9
4 1 3 5 8
5 0 1 3 4
6 0 2 7 8 9
7 0 2 3 6 9
8 2 4 6
9 1 2 3 6 7
```

Evaluation: I had many problems with the assignment. Most I managed to solve, but the one I was left with was converting all the rows, ‘0 1’, ‘0 2’, ‘0 4’ into a single row ‘(0) (1,2,4)’. In the end I was unable to solve this issue and instead made my own data file that reflects how the data should have looked after the first transformation.

From that point it was easy, I sorted the keys, then had to create a function to perform a true mathematical set union. Or taking two sets and creating a new set that consists of the items found in both sets. From there it was just another map reduce problem.

Conclusion: This problem was quite tricky, and provided some challenges that I thought would be simple at first, but I was unable to solve in the time I had. However, by pre-formatting the data in the proper way I was able to solve the problem still.

```
('0 1', {'5', '2'})
('0 2', {'7', '1', '6'})
('0 4', {'1', '2', '4', '5', '6', '7'})
('0 5', {'4', '1'})
('0 6', {'7', '2'})
('0 7', {'6', '2'})
('1 2', {'0', '9'})
('1 3', {'9', '5'})
('1 5', {'0', '3'})
('1 9', {'3', '2'})
('2 6', {'7', '0', '8', '9'})
('2 7', {'0', '9', '6'})
('2 8', {'6'})
('2 9', {'7', '1', '6'})
('3 4', {'1', '5'})
('3 5', {'4', '1'})
('3 7', {'9'})
('3 9', {'7', '1'})
('1 4', {'1', '3', '5', '8'})
('4 5', {'1', '3'})
('4 8', set())
('6 7', {'0', '9', '2'})
('6 8', {'2'})
('6 9', {'7', '2'})
('7 9', {'2', '6', '3'})
```